IMPROVING THE ESTIMATION OF EXTREME SEA LEVELS BY A CHARACTERIZATION OF THE DEPENDENCE OF SKEW SURGES ON HIGH TIDAL LEVELS

Xavier Kergadallan1,2, Pietro Bernardara2,3, Michel Benoit2,4, Camille Daubord5

The knowledge of the statistical distribution of extreme sea levels at the coast is of utmost importance for the characterization of flood risks in coastal areas. In this study we consider that the sea level results from two components: the (astronomical) tide and the (meteorological) surge, without considering the effects of waves. We focus our attention on the dependence of the surge height on the tidal level. At sites with a strong tidal range, the classical analysis methods rely on working only with high tide data (namely high tidal levels and skew surges). A statistical method of adjustment of extreme values is applied to the surge component, leading to the Revisited Joint Probability Method. In that case, we consider that surge and tide components are independent. However, comparisons with measured data show that in several cases this procedure leads to an overestimation of the water levels for a given return period.

We therefore propose here to study the dependence of skew surges on high tidal levels, with two different approaches: one based on a so-called seasonal dependence, and the other one based on the interaction between surge and tide.

Three methods are adapted or developed to test the influence of these two forms of dependence. They are applied to a series of 19 French harbours along the Atlantic and English Channel coasts of France for which more than 10 years of data are available. The results show that the seasonal dependence does not affect the result significantly, while the interaction between the skew surge and the high tidal level appear to be significant for over half the harbours studied.

A revisited model proposed here, as an extension of the model by Dixon and Tawn (1994), seems to be more satisfactory at least for most harbours studied.

Keywords: extreme sea levels, tide surge dependence, skew surge, high tidal range

1 INTRODUCTION

The knowledge of the statistical distribution of sea level at the coast is of utmost importance for the characterization of flood risks in coastal areas, with a particular interest devoted to extreme water levels as they may induce the most dramatic consequences.

In this study we consider and work with the Still Water Level (SWL) which is considered to be the result of two components: tide and (meteorological) surge. The effects of waves are not taken into account here. The classical and widely used method to calculate the distribution of extreme SWL is the Revised Joint Probability Method (RJPM) of Tawn and Vassie (1989). In this approach, the tide and the surge were considered as independent variables.

Some modifications of the RJPM were later proposed by Tawn (1988) and then Dixon and Tawn (1994) to take into account the interaction between surge and tide. The tidal range was divided into bands of equal probability and the distributions of surges were calculated for each tidal band. The influence of the tide on the surge can be explained by the fact that the value of surge is influenced by the water depth for low to moderate water depths (Pugh and Vassie, 1980).

Another type of dependence between surge and tide has been studied by Battistin and Canestrelli (2006). It’s a so called seasonal dependence due to variation of the distributions of surge and tide during the year. Generally the surges are higher between October and March whereas the tidal levels are higher near the equinoxes. So, the higher values of surge and the higher values of tide tend to occur at different periods of the year, and not simultaneously.

In this study, we are more interested into the English Channel and the Atlantic French coasts. That coastline is characterized by quite strong tidal ranges (up to 13 to 14 m at certain locations). Because of the risk to introduce an artificial surge due to uncertainties of temporal phasing on the past SWL observations, a classical way to proceed is to work with high tide water levels (high tidal levels and skew surges) instead of hourly data (e.g. Simon, 1996). The skew surge is thus defined as the difference in sea elevation between the maximum observed water level in a tidal cycle and that predicted maximum (due to tidal effects only).

The latest study of the extreme SWL (again without any effect of waves) along the French coastline of the English Channel and the Atlantic Ocean was provided by SHOM and CETM Ef (2012) (the CETMEF is now part of the Cerema since the beginning of 2014). The method used in that study was the RJPM applied to high tide data (high tidal levels and skew surges). The results tend to show that there are some overestimations of the return level of SWL for some harbours.

1 Cerema Technical Division for Water and Waterways, Plouzané, France
2 Université Paris-Est, Saint-Venant Hydraulics Laboratory (ENPC, EDF R&D, Cerema), Chatou, France
3 EDF Energy R&D UK Centre, London, UK
4 EDF R&D Laboratoire National d’Hydraulique et Environnement (LNHE), Chatou, France
5 SHOM, Brest, France
In the present study we propose to explore the dependence between skew surge and high tidal level to explain the apparent overestimation of the return levels (Kergadallan, 2013). As already mentioned, the effects of waves are not included and the considered water level is the SWL.

The aim of this paper is to examine two different approaches for taking into account the dependence between high tidal level and skew surge: (i) a new “seasonal dependence” approach, and (ii) the original version of the model by Dixon and Tawn (1994) and a newly proposed revisited version to describe the tide surge interaction.

The organization of the paper is as follows. Section 2 describes the different methods used to combine the surge and the tide at high tide to estimate the probabilities of extreme values of SWL. Section 3 illustrates the application of these methods to 19 French harbours along the Atlantic and the English Channel coasts for which more than 10 years of data are available. A discussion of results is provided in Section 4. Conclusions are given in Section 5.

2 METHODOLOGY

2.1 Study area

The study area is limited to the French coasts of the English Channel and the Atlantic Ocean. In the same way as SHOM and CETMEF (2012), it was decided to work with the observations from tide gauges usually located inside harbours. We consider that the quality of the estimations of the extreme sea levels will be too poor for sites with less than 10 years of observations (Simon, 1996), so only stations with at least 10 years of data are selected. The 19 harbours presented in figure 1 fulfil that condition.

2.2 RJPM approach based on the hypothesis of tide surge independence

In the traditional method (JPM), the probability that the still water level \( Z \) exceeds the value \( z \) is estimated by the convolution of surge with tidal level (e.g. Pugh and Vassie, 1979). The tide and the surge are assumed to be independent.

The tail of the surge distribution is estimated by the extreme value theory (Tawn and Vassie, 1989). For that we use a Peak Over Threshold (POT) method with the Generalized Pareto Distribution.
So the probability that the surge $S$ exceeds a given value $s$ higher than the threshold $u$ is defined as follows (Coles, 2001):

$$Pr\{S > s\} = \zeta u \left[ 1 + \xi \left( \frac{s - u}{\sigma} \right) \right]^{-1/\xi}$$

(1)

with $s > u$

$$\zeta_u = Pr\{S > u\}$$

$\sigma$ and $\xi$ are respectively the scale and the shape parameters of the GPD.

The distribution of surges below the threshold is estimated by the empirical distribution. We assume that successive skew surges are independent (Simon, 1996; Bordot and Tawn, 1997). So the empirical part of the surge distribution is directly connected to extreme part.

The probability that the high tidal level $T$ is equal to $t$ is estimated by the empirical function of density of probability with 18 years of predicted values.

So the probability that the SWL $Z$ exceeds the value $z$ is obtained as:

$$Pr\{Z > z\} = \int Pr\{T = t\}Pr\{S > s | s = z - t\}dt$$

(2)

### 2.3 Seasonal dependence approach

During the year the distributions of skew surge and high tidal level are quite different, as illustrated in figure 2 for two harbours (Dunkerque and Le Conquet). The neap tides occur in March and September, whereas the highest surges are associated with storms, which are more intense in winter (say, October to March). This is the so-called seasonal dependence.

We propose a new method to model this seasonal dependence. We consider here there is no physical interaction between the tidal level and the skew surge, but we aim at including the fact that the distributions of skew surge and high tidal level change during the year.

We calculate for each of the 365 days of the year two distributions: one for the high tidal level and one for skew surges. The numbers of days of leap years are normalized at 365 to have always the same number of days per year.

$$Pr\{T_i = t\},$$ the probability that the high tide level $T_i$ of the $i^{th}$ day of the year is equal to $t$, is estimated by the empirical probability density function of high tidal levels.

$$Pr\{S_i > s\},$$ the probability that the skew surge $S_i$ of the $i^{th}$ day of the year is higher than $s$, is estimated by the extreme value theory. In the extreme analysis with the POT method, the probability that the skew surge $S$ exceeds a given value $s$ higher than the threshold $u$ is defined by equation (1).
So \( Pr\{S_i > s\} \) for the \( i^{th} \) day of the year is given by:

\[
Pr\{S_i > s\} = \zeta_{ui} \left[ 1 + \zeta \left( \frac{s-u}{\sigma} \right) \right]^{-1/\zeta} \quad \text{(3)}
\]

with: \( S_i \) the surge of the \( i^{th} \) day of each year,
\( i = 1, \ldots, 365 \) the number of the day in the year,
\( \zeta_{ui} = Pr\{S_i > u\} \) the probability of exceeding \( u \) the \( i^{th} \) day of the year.

The accuracy of the estimation of \( \zeta_{ui} \) is highly dependent on the number of skew surges in the \( i^{th} \) day of the year. This is the reason why we apply a weighted smoothed average of \( \zeta_{ui} \) on 31 days:

\[
\zeta_{ui} = \frac{1}{31} \sum_{i=-15}^{+15} (16-i) Pr\{S_j > s\} \quad \text{(4)}
\]

with:
- \( j = i - 365 \) if \( i < 1 \)
- \( j = i \) if \( 1 \leq i \leq 365 \)
- \( j = 365 - i \) if \( i > 365 \)

So, by including the seasonal dependence in this way, the probability that the SWL \( Z \) exceeds the value \( z \) is given by:

\[
Pr\{Z > z\} = \sum_{i=1}^{365} Pr\{T_i = t\} Pr\{S_i > s\} s = z - t dt \quad \text{(5)}
\]

### 2.4 Tide surge interaction models

#### 2.4.1 Dixon & Tawn (1994) model

The characteristics of the interaction between the surge and the tidal level vary from site to site due mainly to variations in water depth. Dixon and Tawn (1994) proposed a model to describe the interaction between the extreme instantaneous surges and the associated tidal level. The model is transposed here to the interaction between the extreme skew surges and the associated high tidal level.

The upper tail of the surge distribution \( S \) is modelled conditional on the tide level \( T \) with introduction of the variable \( S^* \) the surge normalized per 5 or 10 equi-probable tidal bands:

\[
S^*(T) = \frac{s-a(t)}{b(t)} \quad \text{(6)}
\]

with \( a(T) \) and \( b(T) \) estimated from piecewise linear functions of the 98% and 99% empirical quantiles of measured skew surges in each tidal band (see figure 3 for two harbours, Dunkerque and Le Conquet).

![Figure 3: 98% and 99% empirical quantiles of skew surges in each of the 5 tidal bands of equal probability.](image)
In equation (6) the values of $a(t)$ and $b(t)$ are respectively $u_i(t)$ and $u_j(t) - u_i(t)$, defined as follows:

- For Middals bands: $[t_{01}, t_{11}]$, $[t_{11}, t_{21}]$, ..., $[t_{N,1}, t_{N1}]$,
- and $m_i$ the mid-point of the $i$th tidal band : $m_i = (t_{i,1} + t_{i,2})/2$,
- with $q_{i,1}$ and $q_{i,2}$ respectively the quantiles 98% and 99% of the empirical distribution of skew surge in the $i$th tidal band,
- $u_i(t)$ is given by

$$
\begin{align*}
q_{j,1} & \quad \text{for } t_0 \leq T < m_1 \\
q_{j,i} & = \frac{T - m_{i-1}}{m_i - m_{i-1}} + \frac{T - m_i}{m_{i-1} - m_i} q_{j,i-1} \\
& \quad \text{for } m_{i-1} \leq T < m_i \text{ and } i = 2, \ldots, N \\
q_{j,N} & \quad \text{for } m_N \leq T \leq t_N
\end{align*}
$$

This model is not directly applicable to the empirical part of the skew surge distribution (skew surges below selected threshold). That is why we work here with only the extreme skew surges (i.e. the upper tail of the distribution).

A POT method is applied to the normalized surge. So the probability that the SWL $Z$ exceeds the value $c$ is given by equation (2) with $Pr[S > s]$ function of $t$:

$$
Pr\{S > s\} = Pr\{S^* > s^* \mid s^* = \frac{s - a(t)}{c(t)}\}
$$

(8)

The Dixon and Tawn method was initially applied to instantaneous surge and the whole tidal range, with 5 to 10 tidal bands. In the present study, the Dixon and Tawn method is applied to skew surge and high tidal level with 5 tidal bands. Due to the weaker range of tidal levels for a high tide study in comparison to the whole tidal range, the discretization in tide is finer here than in Dixon and Tawn (1994).

2.4.2 Revisited Dixon and Tawn model

Because we want to test the sensitivity of the model to the number of equi-probable tidal bands we propose to modify the method. If we increase the number of equi-probable tidal bands, the number of elementary events per band decreases, so the accuracy of the quantiles 98% and 99% of the empirical distribution of skew surges in the $i$th tidal band decreases. To normalize the skew surges per band we choose a method close to the one used in the regional frequency analysis method (Bernardara et al., 2011) with $S'$ the normalized surge:

$$
S'(T) = \frac{s}{c(t)}
$$

(9)

In equation (9) the value of $c(t)$ is defined as follows (see figure 4):

- for $N$ tides bands : $[t_{01}, t_{11}]$, $[t_{11}, t_{21}]$, ..., $[t_{N,1}, t_{N1}]$,
- and $m_i$ the mid-point of the $i$th tidal band : $m_i = (t_{i,1} + t_{i,2})/2$,
- with $q_{av,i}$ the mean of surge in the $i$th tidal band exceeding the quantile 95% of the empirical distribution of skew surges in the $i$th tidal band,
- $c(t)$ is given by:

$$
\begin{align*}
q_{av,1} & \quad \text{for } t_0 \leq T < m_1 \\
q_{av,i} & = \frac{T - m_{i-1}}{m_i - m_{i-1}} + \frac{T - m_i}{m_{i-1} - m_i} q_{av,i-1} \\
& \quad \text{for } m_{i-1} \leq T < m_i \text{ and } i = 2, \ldots, N \\
q_{av,N} & \quad \text{for } m_N \leq T \leq t_N
\end{align*}
$$

(10)
Similar to the Dixon and Tawn method, the model is not directly applicable to the empirical part of the skew surge distribution. Because we want to compare the estimations of SWL with the observations of SWL, we simplify the model to get values in the non-extreme part of the estimated distribution.

\[
S'(T) = \frac{s}{c(t)} \quad \text{with} \quad c(t) = q_{av} \quad \text{for} \quad t_{i-1} \leq T < t_i
\]  

We control for the number of tidal bands used here that this simplification does not modify the result significantly. So with this new model we can easily extend the estimated distribution of skew surges to values below the threshold. To do that, we calculate the empirical distribution of skew surges for each tidal band.

A POT method is applied to the new normalized surge. So the probability that the SWL \( Z \) exceeds the value \( z \) is given by equation (2) with \( Pr\{S > s\} \) function of \( t \):

\[
Pr\{S > s\} = Pr\{S' > s' \mid s' = s/c(t)\}
\]  

In this revisited method we test the influence of the number of tidal bands. A high number of tidal bands is expected to result in the most accurate model of tide surge interaction if however we have enough events per band. After different tests we chose to relate the number of tidal bands to the duration of observations, by adopting the following rule: defining \( X \) tidal bands for \( X \) years of observations.

3 APPLICATIONS

3.1 Results

For each harbour the return periods of SWL are computed in 5 different ways:

- 1 direct method: the empirical frequency of sample value (Weibull plotting position), called, “empirical model”
- 4 indirect methods with the convolution of skew surge with high tidal level:
  - RJPM, based on a hypothesis of independence between tide and surge,
  - a model of seasonal dependence, called “seasonal model”
  - a model of tide surge interaction with Dixon and Tawn method, called “Dixon & Tawn model”
  - a model of tide surge interaction with recasting method, called “revised model”.

The results are presented in figures 5 to 8 for four selected harbours. These harbours were chosen because they are representative of all the other ones. Because it does not benefit to the discussion and to improve the clarity of the figures, the confidence intervals are not drawn. The confidence intervals could be easily calculated by a Monte-Carlo method.
3.2 Analysis

The goodness-of-fit of the four indirect methods is appreciated through visual inspection of figures 5 to 8. We consider that one model is better than the others if the direction of its upper tail of distribution is more similar to the direction of the empirical distribution. We consider that the estimations of SWL resulting from two different models are significantly different if the absolute difference of the values at the 100 years return period is greater than 5 cm. We choose this threshold of 5 cm because it is the order of accuracy of mechanic tide gauge measurement.

The results provided from “seasonal model” are significantly close to those from RJPM (compare blue and cyan curves in figures 5 to 8). The absolute differences of the 100-year SWL return levels between “seasonal model” and RJPM are less than 5 cm for all of the 19 harbours.

The absolute difference of the 100-year SWL return levels between the “Dixon & Tawn model” and RJPM is not significant for 6 harbours (see an example of these 6 harbours, Calais, in figure 5). We consider the goodness-of-fit with empirical data of the “Dixon & Tawn model” is better than for RJPM for 7 harbours (see an example, Arcachon, in figure 6). The results with “Dixon & Tawn model” would be unsatisfactory for 6 harbours (see examples in figures 7 and 8).

The absolute difference of the 100-year SWL return levels between the “Revisited model” and RJPM is not significant for 5 harbours (see figure 5). We consider the goodness-of-fit with empirical data of the “Revisited model” is better than for RJPM for 13 harbours (see figure 7). The results with the “Revisited model” would be unsatisfactory for only 1 harbours (see figure 8).

If we compare only the goodness-of-fit of the “Dixon & Tawn model” with the “Revisited model”, the difference is not significant for 6 harbours. Among the remaining 13 harbours, the “Revisited
model” would be more efficient for more than half of the harbours (9 harbours in favour of the “Revisited model”).

These results thus tend to show that the seasonal dependence does not impact significantly the estimation of SWL return level. If interaction between the high tidal level and the skew surge are taken into account, the results for around 70% of the harbours are modified. So for 30% of the harbours the skew surge and the high tidal level could be considered as independent. To model the tide surge interaction, the revisited model appears to be more efficient than the “Dixon & Tawn model” in most of cases, but not systematically.

4 DISCUSSION
The goodness-of-fit was appreciated though a visual inspection in the previous section, although that way to analyze the tested models is not obvious. Because the duration of observation is important for the highest values of observations. In addition the choice of the formula to calculate the empirical return period (“Weibull plotting position” here) modifies the result for these same high values (see, e.g., Makkonen (2006) and Cook (2011) for a discussion about the plotting position). However if we set aside these highest values the shape of the empirical distribution remains instructive to get an idea of the goodness-of-fit.

An alternative way would be to quantify the goodness-of-fit with some statistical tests. We have tried different statistical tests, namely Chi2, Kolmogorov-Smirnov, Anderson-Darling and modified Anderson-Darling tests. The problem is that the results depend on the choice of the test and the parameters of that test (the threshold used to select the population for example). So the choice of one test appears to introduce a too strong influence on the analysis.

The visual analysis is more complicated for the “Dixon & Tawn model”. That model is not defined on the empirical part of the distribution. So the estimations begin at the 10-year or 100-year return period. So it is very difficult to judge the goodness-of-fit to the data. It would be useful to have some results from the “Dixon & Tawn model” for lower return periods.

The results show that the tide surge interaction at high tide cannot be neglected. But even if the “Revisited model” seems to perform slightly better than the “Dixon & Tawn model”, none of these models is completely satisfactory. The results between neighbouring harbours are not always consistent while a physical criterion in favour of one model would be to find a tendency on large areas.

Finally in all this work we have neglected the dependence between successive values. We consider that two successive high tidal levels or skew surges are independent. But we know that the duration of storm is often higher than 2 or 3 days. That hypothesis of independence could change the results of the tested models. Its influence has thus to be further tested.

5 CONCLUSIONS AND OUTLOOK
Traditionally at sites with a strong tidal range the still water return levels are estimated by a method of convolution of skew surge distribution with the distribution of predicted high tidal levels, through the Revised Joint Probability Method of Tawn and Vassie (1989). The current part of the distribution of skew surges is defined by an empirical way and the upper tail is defined by an analysis of the extreme values. This approach assumes that skew surges and high tidal levels are independent.

However when we compare estimations from the RJPM with the empirical probabilities of observations of SWL, one might wonder if the classical method overestimates or not the return levels. One explication could be that skew surges and the high tidal levels are dependent.

Two different types of dependence could affect the result:

- a seasonal dependence because high values of surges and tidal levels are not expected to occur during the same period of the year.
- a physical interaction between surge and tide due to the variation of the water depth.

To check the validity of these hypotheses three models are tested: a new model for seasonal dependence, an existing model the “Dixon and Tawn model” for interaction between surge and tide and for the last test some modifications were made to the “Dixon and Tawn model”, giving the “Revisited model”.

The results show that the seasonal dependence does not affect the estimations significantly, while the interaction between the skew surge and the high tidal level could be significant for over one half of the harbours studied. The “Revisited model” appears to be the more satisfactory method to capture that interaction. But the lack of spatial coherence of the results tends to indicate that the model needs to be improved.
Another question would be the validity of the hypothesis of independence between successive values of skew surge and high tide level. It would thus be interesting to extend that study with a model of dependence between successive values.

6 ACKNOWLEDGEMENTS

Tide gauges observations of Arcachon, Boulogne sur Mer, Brest, Bayonne, Calais, Cherbourg, Concarneau, Dieppe, Dunkerque, Le Havre, Le Conquet, La Rochelle, Port-Bloc, Port-Tudy, Roscoff, Saint Jean de Luz, Saint-Malo, Saint-Nazaire, Les Sables d'Olonne are the property of SHOM, Marine Nationale, GPM Dunkerque, GPM du Havre, GPM de Bordeaux, GPM de Nantes-Saint Nazaire, GPM La Rochelle-La Pallice, Port de Calais, Port de Boulogne-sur-Mer, Port de Dieppe, Port de Cherbourg, Région Nord Pas de Calais, CR Aquitaine, CG du Finistère, CG Pyrénées Atlantiques, CG du Morbihan, CCI pays de Saint-Malo, CCI de Morlaix, CCI Quimper Cornouaille, CCI de Vendée, DDTM Ille-et-Vilaine, DDTM de Gironde and Mairie de Groix.

Tide gauge observations are available on the REFMAR website (http://www.refmar.shom.fr).

Predicted tidal levels are provided by SHOM (bp@shom.fr).

7 REFERENCES


SHOM and CETMEF. 2012. Statistiques des niveaux marins extrêmes des côtes de France (Manche et Atlantique), Édition SHOM, Brest (France) (in French).
